

# Naturalment

## Entrevista

Roderic Guigó



origen: [www.20minutos.es](http://www.20minutos.es)

Doctorat per la Universitat de Barcelona (1988), va treballar al departament de Bioinformàtica del Dana Farber Cancer Institute (Harvard) i a Los Alamos National Laboratory. El 1994 va començar a treballar a l'Institut Municipal d'Investigació Mèdica. Actualment és catedràtic de bioinformàtica a la Universitat Pompeu Fabra. A més és autor de més de 120 publicacions i és col·laborador del projecte ENCODE. (informació consultada a: <http://www.cicbiogune.es/>)

### En què consisteix el projecte ENCODE?

ENCODE vol dir ENCyclopedia Of DNA Elements. El projecte consisteix precisament a identificar tots els elements de la seqüència del genoma humà que confereixen funcionalitat. Essencialment aquelles regions que es transcriuen a RNA (els gens) i aquelles regions que controlen aquesta transcripció, ja siguin proximals als gens (com els promotors) o distals (com els enhancers). També, es clar, altres regions la funcionalitat de les quals no està directament relacionada amb la transcripció, com ara llocs d'inici de la replicació, regions d'interacció entre el DNA, etc. Per tal d'identificar aquestes regions el projecte ENCODE utilitza una combinació de mètodes experimentals (majoritàriament basats en la seqüència massiva: RNASeq, ChIPSeq, CLIPSeq etc) i computacionals.

### Per què vareu començar el projecte ENCODE?

Un cop publicat el primer esborrany de la seqüència del genoma humà l'any 2001, els instituts nacionals de la salut (NIH, en anglès) que van ser els que van finançar majoritàriament el projecte del genoma humà, es van plantejar quins haurien de ser les passes següents per tal de treure el màxim profit d'aquesta seqüència. Aleshores van decidir finançar projectes en dues direccions complementàries. D'una banda, l'esborrany de la seqüència que es va publicar l'any 2001 corresponia a una única seqüència—diguem-ne que a una mena de seqüència arquetípica—però no corresponia a la seqüència de ningú en concret. Però nosaltres sabem que les diferències biològiques entre els individus es deuen majoritàriament a diferències en la seqüència del genoma. Aleshores els NIH van iniciar un projecte o projectes encaminats a caracteritzar la variabilitat genòmica dins de l'espècie humana. El projecte inicial es va anomenar HapMap, i consistia en

catalogar la variació humana només en unes posicions específiques del genoma, però a mesura que les tecnologies de seqüenciació van continuar progressant es va plantejar el projecte de seqüenciar el genoma complet de molts individus humans. Així va començar el projecte dels 1000 genomes, els resultats del qual s'han donat a conèixer fa unes poques setmanes. La segona línia de recerca que l'NIH va considerar prioritària va ser la de la investigació del funcionament del genoma. El genoma es defineix sovint com el conjunt d'instruccions que determinen les característiques biològiques dels éssers vius, però com es produeix la codificació d'aquestes instruccions en el genoma ens es majoritàriament desconegut. Per tal d'entendre com es produeix aquesta codificació—es a dir, d'identificar aquelles regions del genoma que són funcionals, que codifiquen informació biològica—es va iniciar el projecte ENCODE

### **Quants països estan involucrats amb el projecte i quina implicació té l'estat espanyol?**

El projecte ENCODE es un projecte bàsicament americà. És finançat al 100% pels Estats Units (NIH). Contràriament, però, al que passa a la majoria de països europeus, on les agències de recerca nacionals només subvencionen recerca dins el país, en principi qualsevol investigador del món pot presentar un projecte al NIH, i rebre subvenció. La subvenció d'investigadors estrangers no és la norma, però tampoc es excepcional. En el cas concret d'ENCODE, es van subvencionar una desena de subprojectes. Cada un d'aquests projectes té un investigador principal (PI). Normalment els PIs són americans, i els investigadors estrangers s'integren als equips com a co-investigadors. A la fase actual d'ENCODE només hi ha hagut un subprojecte en el qual l'investigador principal no fos americà. Ha estat liderat per Tim Hubbard del Sanger Institute al Regne Unit. Jo sóc co-PI en aquest projecte, i en el projecte liderat pel Tom Gingeras de Cold Spring Harbor (a l'estat de Nova York). [Jo vaig ser PI d'un projecte a la fase pilot d'ENCODE]. A banda dels grups americans, del Regne Unit i de l'Estat Espanyol, també han participat en el projecte ENCODE, investigadors de Suïssa, Singapur i Japó. A l'Estat Espanyol, la participació més important ha estat la del nostre grup al CRG, donat que hem liderat el grup d'anàlisi del RNA. També hi ha hagut, però, participació del grup de l'Alfonso Valencia al Centro Nacional de Investigaciones Oncológicas (CNIO)

### **El concepte de genoma canviarà o es definirà de manera diferent de com ha estat definit fins ara? - Significa que l'ADN "fems" no existeix?**

El concepte bàsic del genoma no canviarà. El genoma es el conjunt d'instruccions que determinen les característiques biològiques dels éssers vius i aquest concepte, com a tal, no canviarà. El que passa és que el genoma "funciona" de manera diferent en tipus cel·lulars diferents. Es per això que cèl·lules amb exactament el mateix genoma dins d'un organisme són tant diferents (com les cèl·lules de la pell, dels ossos, les neurones, les cèl·lules musculars, etc). Que el genoma funciona de manera diferent a tipus cel·lulars diferents simplement vol dir que el RNA que el genoma produeix es diferent en aquests tipus cel·lular. Inicialment es pensava que només una part petita del genoma estava dedicada a la producció de RNA (i la resta s'anomenava DNA escombraria/deixalla/fem/etc "junk DNA"). El projecte ENCODE ha revelat que fins el 60% de la seqüència del genoma es capaç de produir RNA. A més, una gran part del 40% restant podria estar involucrada en la regulació diferencial de la producció de RNA. Es a dir, s'hi trobarien les regions que controlen quin RNA es produeix específicament (és a dir, quins gens s'expressen, funcionen) en un determinat tipus cel·lular. Globalment, el projecte ENCODE ha detectat activitat en el 80% de la seqüència del genoma. Però activitat no vol dir necessàriament funció biològica. Es a dir, és possible que es produeixin molècules de RNA que finalment no tinguin cap paper biològic. Determinar quina fracció d'aquest 80% del genoma que sembla actiu és realment funcional, és la tasca de la biologia dels propers anys i dècades.